Salient object detection employing regional principal color and texture cues

Mudassir Rafi[1] · Susanta Mukhopadhyay[1]

Abstract

Saliency in a scene describes those facets of any stimulus that makes it stand out from the masses. Saliency detection has attracted numerous algorithms in recent past and proved to be an important aspect in object recognition, image compression, classification and retrieval tasks. The present method makes two complementary saliency maps namely color and texture. The method employs superpixel segmentation using *Simple Linear Iterative Clustering (SLIC)*. The tiny regions obtained are further clustered on the basis of homogeneity using DBSCAN. The method also employs two levels of quantization of color that makes the saliency computation easier. Basically, it is an adaptation to the property of the human visual system by which it discards the less frequent colors in detecting the salient objects. Furthermore, color saliency map is computed using the center surround principle. For texture saliency map, Gabor filter is employed as it is proved to be one of the appropriate mechanisms for texture characterization. Finally, the color and texture saliency maps are combined in a non-linear manner to obtain the final saliency map. The experimental results along with the performance measures have established the efficacy of the proposed method.

1 Introduction

In a scene, the human visual system (HVS) does not attend all the regions uniformly. It is heavily biased towards certain objects and such behavior is persistent for all the subjects. These preferential objects are known as salient objects within the scene. Similarly, the human visual system (HVS) is also biased towards certain spatial regions or objects in the image as well. These spatial regions or objects are known as salient regions or salient objects within the image. The salient objects do not depend upon the purpose of the image

✉ Mudassir Rafi
mudassir.rafi23@gmail.com

Susanta Mukhopadhyay
msushanta2001@gmail.com

[1] Indian Institute of Technology (ISM), Dhanbad, India

analysis instead the salient objects are determined in the context of the human visual system (HVS). The detection of these preferential objects is known as as saliency detection. Moreover, the formation of gray scale maps of these preferential objects are known as saliency maps. The task of salient object detection includes two important steps : (a) Detection of the most salient object (b) Segmentation of the accurate regions of these objects. Salient object detection should not be confused with image segmentation as image segmentation is a part of the salient object detection process. The present paper addresses the problem of ascertaining the saliencies to the objects in the form of grayscale saliency maps which can further be segmented out using Otsu's thresholding [35] for salient object segmentations. Salient object detection models are important because of their wide applications in the areas of computer vision, graphics and robotics. In recent past, the saliency models have been employed for image segmentation [5, 12], image classification [41, 43], object recognition [40], image retrieval [6, 14], image and video compression [11], video summarization [20], image quality assesment [26], visual tracking [27], non-photorealistic rendering [8, 18] and human-robot interaction [33]. Saliency detection is strictly related to human perception of visual information; thereby, remains a subject matter for various disciplines comprising cognitive psychology, neurobiology, and computer vision.

The main objective of the saliency detection is to find those regions which are in agreement with the manually annotated ground truth. In saliency detection algorithms, feature extraction plays a key role. The existing algorithms can mainly be classified into two categories. First category of methods employs color as the principal feature in order to compute saliency maps [1, 2, 8, 10, 15, 19, 27, 32]. These methods are unable to handle complex texture images. The second category consists those methods which use both color and texture [7, 52]. However, a very few methods come under this category, as the use of textural features for saliency detection has not been explored much. In this category, some of the methods have proposed adaptive fusion strategy for combining the two feature maps i.e. color and texture. In order to compute the adaptive fusion, they have proposed the computation of the optimum fusion ratio of the color and texture. To the contrary, the adaptive combination does not always yield good results. Therefore, the method presented in this paper does not employ the adaptive fusion strategy, rather it uses a linear combination approach, so that the contribution of the two maps can easily be varied, wherever necessary. For texture map, the present method employs directional filters which not only produce a prominent texture map for prominent texture images but also contribute enough in the final saliency for objects with prominent orientation (as shown in Fig. 1d).

In case of color saliency, a regional principal color contrast is proposed which employs color features to detect saliency on the same scale as of the input image. In due course of color saliency computation, the method employs the region pair contrast measurement [31] which is found to be more efficient than the pixel pair contrast. The method starts with the quantization of colors which quantizes further using color histograms. Furthermore, the
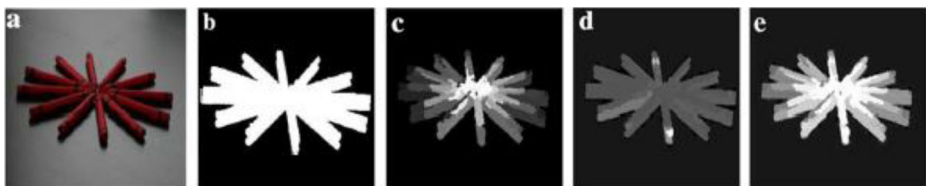
Fig. 1 **a** Input Image **b** Ground Truth **c** Color Saliency Map **d** Texture Saliency Map **e** Final Saliency map

quantized image is segmented into regions and the principal color of each region which is employed for saliency computation. Moreover, two types of spatial relationships are also computed and employed with regional principal color. Again, the two saliency maps namely color and texture are also combined together using linear combination to form final saliency map. Then, the present method is compared with [1, 2, 10, 15, 19, 22, 32, 51].

The major motivation to the proposed method comes from the fact that, in real life, most of the time we are encountered with textural objects in addition to the color. Additionally, texture is considered as a primary low level cue along with color for object recognition and identification tasks. Thus, it is not justified to ignore one of the important recognition features for saliency detection. As in [31], authors have completely ignore texture. Thus, authors have proposed a modified method based on [31] which also takes into consideration texture features, extracted using Gabor filter (Fig. 2).

The remainder of the paper is organised as follows: Section 2 presents a brief overview of similar works related to visual saliency detection, Section 3 describes the Gabor Filter used in the present work. Section 4 describes the super pixel segmentation employing SLIC. Section 5 describes the proposed method which has further been subdivided into Color and Texture saliency maps. Section 6 presents the experimental prerequisites, procedure and a comprehensive discussions on the output results. The paper ends with a conclusive summary.

2 Related work

In literature, various methods for visual saliency have been proposed. The earlier methods exploit features extracted from blocks. One of the most acclaimed methods under this category is proposed by Itti et al. [19]. They devised an algorithm on the basis of the behavior of visual receptive field, and employed the centre surround differences. Along the same line, Hu et al. [17] used polar transformation for representing an image in 2D space and mapped them to 1D subspace. Subsequently, Generalized Principal Component Analysis (GPCA) is applied where salient regions are computed using feature contrasts and regional geometric properties. Similarly, Rosin et al. [39] proposed an efficient and parameter free approach whereas Valenti et al. [46] linearly combined the curvedness, color boosting and iso-centre clustering. On the other hand, Achanta et al. [2] proposed a frequency tuned filter to compute pixel saliency. Moreover, Frintrop et al. [24] used an information theoretic approach to compute saliency using Kullback-Leibler divergence. Wang et al. [48] suggested a saliency method based on non-local reconstruction; the method targeted more on image singularity. Hu et al. [53] proposed a pixel based saliency method assimilating compactness and local contrast using diffusion process. Ma and Zhang [38] introduced a different approach employing local contrast in order to create a saliency map which is extended further
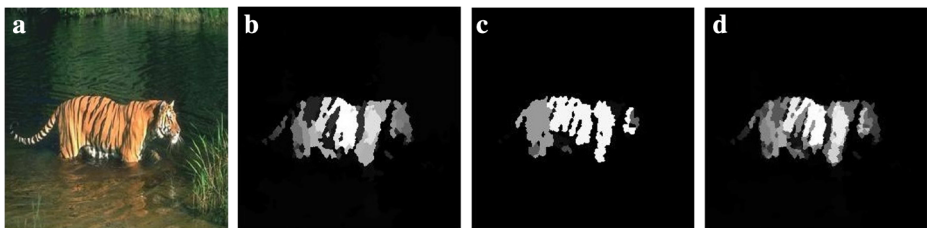
Fig. 2 **a** Input Image **b** Color Saliency Map **c** Texture Saliency Map **d** Final Saliency map

employing a fuzzy growth model. Harel et al. [45] modified the method of Itti et al. [19] by normalizing the feature maps so that discernible parts are highlighted and allowed to integrate with other important maps. Goferman et al. [10] proposed a method for context aware saliency employing low and high level features with global considerations. The proposed saliency is some how different from the earlier proposed saliencies. Li et al. [28] formulated this problem in terms of cost sensitive max-margin classification problem.

The approaches discussed above detect the salient objects based either on pixels or patches. The main deficiency in these approaches are: first, the high contrast edges become prominent instead of salient objects; and second, the boundaries of salient objects are not well localized. Therefore to overcome these issues, region based methods are proposed.

Region based methods, generally, segment an input image into regions aligned with the intensity edges and then compute the regional saliency map. The initial methods under this category have computed a regional saliency score [29]. In Yu et al. [50], it is computed in terms of background score. The background score of each region is computed employing the observations from the background and salient regions. Moreover, there are methods which defined saliency in terms of uniqueness of global region contrasts [49]. Perazzi et al. [36] decomposed the image into homogeneous regions and, handled local and global contrast in a unified manner. Similarly, Chang et al. [4] employed a graphical model framework by fusing the objectness and regional saliency where these two terms are computed collectively by using the energy function. Jiang et al. [21] defined the regional objectness as the average objectness values of the pixels and employed it for regional saliency computations.

The methods described are based on regional saliency computations. These methods are found to be advantageous in comparison to block based methods because their regions are fewer in numbers. the regions are few in numbers. Thus, saliency can efficiently be computed. Additionally, the region based methods lead to the extraction of more informative features as the regions are non-overlapping and homogeneous. Therefore, in the present work authors have employed the region based framework.

3 Gabor filter

Gabor filters are widely used in describing textures. They perform very well in classifying images with different textures. The neuroscientists believe that receptive fields of the human visual system (HVS) can be represented as basis functions similar to Gabor filters [34]. A Gabor function is a gaussian modulated sinusoid [9]. In spatial domain, the gabor filter is represented as a two dimensional impulse given by

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} exp\left(-\frac{1}{2}\left[\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right]\right) exp(j2\pi Fx + \phi) \tag{1}$$

where F and ϕ are the frequency and phase of the sinusoidal wave. The values of σ_x and σ_y are the size of the gaussian envelopes in the x and y directions respectively. Moreover σ_x/σ_y is known as aspect ratio denoted by γ. Correspondingly, in spatial-frequency domain the gabor filer is given by

$$G(u, v) = exp(-2\pi^2[(u - F)^2\sigma_x^2 + v^2\sigma_y^2]) \tag{2}$$

The selection of appropriate parameters for gabor filter has always been important in gabor based image processing. In literature, a great deal of work has been done on this subject in order to restrict the degree of freedom of gabor filters based on neurophysiological findings [37].

In the present work a directional filter bank (Fig. 3) of 40 filters with 5 scales and 8 orientations are used. The filters are separated from each other by 22.5° so that a finer quantization can be obtained.

4 SLIC

Superpixels can be defined as clustering of pixels having homogeneous characteristics. It provides an appropriate way to compute local image features and minimize immensely the intricacy of the image processing tasks. The superpixels have been proved beneficial for depth estimation, image segmentation, skeletonization, body model estimation and object localization. In literature, various method have been employed for superpixel segmentation; but, they suffered from high computational cost, poor quality segmentation, inconsistent size and shape, and a number of parameters which neeed initilaization. SLIC [3] stands for *simple linear iterative clustering* and employs a local clustering of pixels in the 5d space defined by L, a, b values of the CIELAB color space and the spatial co-ordinates x, y. The major advantages of SLIC are: easy to implement, less computational cost and a single parameter in the form of number of desired superpixels. However, it has a limitation; the RGB image has to be converted to CIELAB colorspace which can be ignored to the offered benefits.

The algorithm starts by sampling K in regularly spaced cluster centres and by moving them to the seed locations corresponding to the lowest gradient position in an $m \times m$ neighbourhood (In the present work, $m = 3$). The algorithm not only avoids the placement over edges but also avoids the selection of a noisy pixel. Every pixel in the image is associated with a cluster centre whose area overlaps with this pixel. After all the pixels are associated with the nearest cluster center, a new center is computed as the average $labxy$ vector of all the pixels belonging to the cluster. We then iteratively repeat the process of associating pixels with the nearest cluster center and recompute the cluster center until convergence is reached.

At the end of this process, a few stray labels may be there i.e a few pixels present in the vicinity of a larger segment having the same label but not connected to it. Although, it is rare, it is possible. At last, connectivity is enforced by relabeling the disjoint segments with the labels of the largest neighboring cluster.

(a) Real Part of Gabor filter   (b) Magnitude of Gabor filter
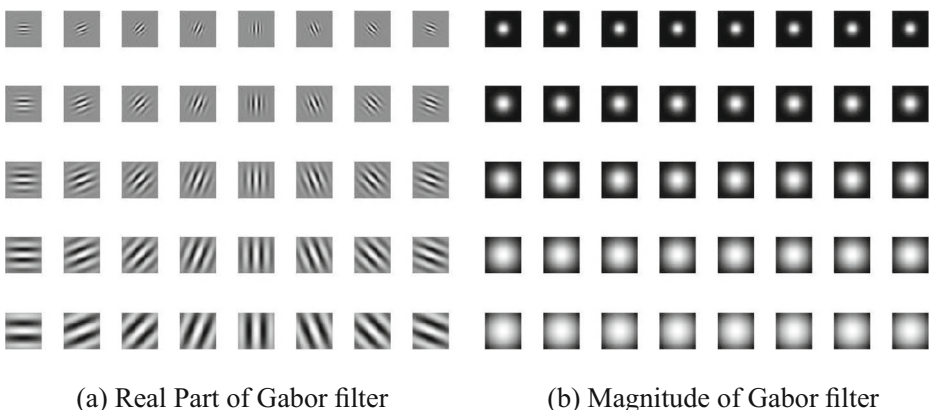
Fig. 3 Gabor filter bank

5 The proposed method

The proposed method is divided into three main parts namely, (a) Formation of Color Saliency Map (b) Formation of Texture Saliency Map and finally, a step which combines both. The present method employ superpixels segmentation, which segments an image into several smaller homogeneous regions. It has been used as a prior step in many computer vision tasks because of its pivotal role in object boundary preservation. Furthermore, the true color RGB image is subjected to *minimum variance quantization* and regional principal colors.

5.1 Formation of Color Saliency Map

In a 24-bit RGB image, there are 256^3 number of colors. Out of these, a handful number of colors are frequently used. The color shades which have minute differences are not only difficult to recognize but most of the time categorizes into one another. Furthermore, these less frequent colors have no contribution in saliency detection and increase computational complexity during saliency computations. The frequent colors are not only crucial but also decisive for salient objects detection. Thus, it is wise to discard the infrequent colors and replace them with the frequent color values. For this reason, the present method (Fig. 4) quantizes the true color image into 256 colors employing minimum variance quantization. The resultant image is subjected to superpixel segmentation, which divides it into a number of regions. For superpixel segmentation, $SLIC$ algorithm has been used, the parameter i.e. number of superpixels is taken as 3000. As mentioned in previous section; $SLIC$ is used because of less computational overhead compared with other algorithms. However, it results into over segmentation. Thus, to make the number of segmented regions manageable and extracting a meaningful information from them, DBSCAN clustering is used. The DBSCAN clustering algorithm is robust to outliers and requires only two parameters. However, it produces homogeneous regions, it does not discard the presence of other colors or color shades in these regions. Therefore, in each region we searched for regional principal color by employing color histogram. The color histogram can be used to know the frequency of frequent occuring color shades and can replace the other colors. After this step, we have superpixels from which saliency is computed.

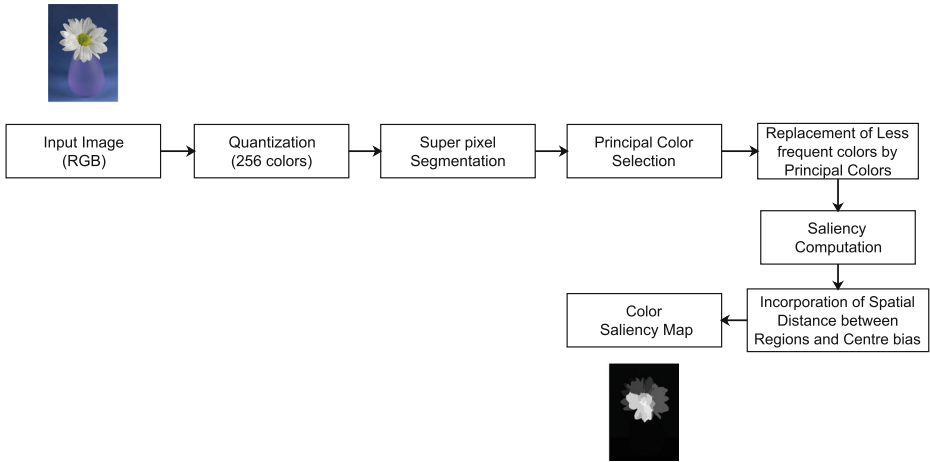
Algorithm 1 Proposed algorithm (Color Sal. map).

1: **procedure** COL–SALMAP(**Input** Image $I_{rgb}(x, y)$)

2: $I_{quant}(x, y) = MVQ(I_{rgb}(x, y))$

3: $I_{superpix}(x, y) = DBSCAN(SLIC(I_{quant}(x, y)))$

4:     **for** each region $r_i(x, y) \in I_{superpix}(x, y)$ **do**

5:         $I_{RPC}(x, y) = RPC(r_i(x, y))$

6:     **end for**

7:     $Sal(c_i) = \sum_{j=1}^{N} nd(c_i, c_j)$

8:     $Sal_{smth}(c_i) = \frac{1}{(p-1)G} \sum_{j=1}^{p} (G - d(c_i, c_j)Sal(c_i))$

9:     $Sal_{reg}(r_i) = n_i Sal_{smth}(r_i) + \sum_{j \neq i} n_j \psi(r_i, r_j) e^{-d(r_i, r_j)^2}$

10:    $Sal_{col}(r_i) = \dfrac{Sal_{reg}(r_i)}{e^{d(r_i, C)^2/\delta}}$

11: **end procedure**

Fig. 4 Color saliency map

With no previous information about the size and location of the salient object, the saliency value of a specific color in a color image can be described mathemaically, in terms of the global color contrast values as:

$$Sal(c_i) = \sum_{j=1}^{N} n d(c_i, c_j) \qquad (3)$$

Here, c_i is the color for which color saliecy has been determined, N denotes the number of colors, n is the total number of pixels having color c_j and d is the euclidean distance between color values c_i and c_j. Furthermore, Cheng et al. [8] replaced the saliency value of a color with p most similar colors in the LAB color space.

$$Sal_{smth}(c_i) = \frac{1}{(p-1)G} \sum_{j=1}^{p} (G - d(c_i, c_j) Sal(c_i)) \qquad (4)$$

where, N is same as in the previous equation, $p = \lceil \delta \times N \rceil$ represents number of similar colors to c_i and $G = \sum_{j=1}^{p} d(c_i, c_j)$ represents sum of color contrasts between c_i and other colors.

As it has already been established that human visual system (HVS) uses a centre surround principle [24] for scanning a scene. Moreover, the attention shifts from one salient region to the next most salient region. Along the same line, researchers [23, 44] have found that in an image the regions at the centre are prone to be more salient than those present at the boundaries. Thus, this concept of centre bias has been mathematically, employed as gaussian distribution by [21, 42, 47, 49]. Thus, the regional saliency [31] of a region considering the distance between two regions

$$Sal_{reg}(s_i) = n_i Sal(r_i) + \sum_{j \neq i} n_j \psi(r_i, r_j) e^{-d(r_i, r_j)^2} \qquad (5)$$

where n_i represents the number of pixels in regions r_i, n_j represents the number of pixels in r_j and $d(r_i, r_j)$ denotes the euclidean distance between regions r_i and r_j normalized to

[0, 1]. Whereas $\psi(r_i, r_j)$ can be defined as

$$\psi(r_i, r_j) = \left\{ \begin{array}{ll} Sal(r_i) - Sal(r_j) & if\, Sal(r_i) > Sal(r_j) \\ 0 & otherwise \end{array} \right\} \quad (6)$$

Further more, color saliency after incorporating the regional and image centre bias

$$Sal_{col}(r_i) = \frac{Sal_{reg}(r_i)}{e^{d(r_i, C)^2/\delta}} \quad (7)$$

Here, $\delta \epsilon(0, 1]$ represents the strength of the response of the spatial weighting.

5.2 Formation of texture saliency map

Texture is a collective repetitive pattern that characterizes the surface of real world objects. It plays an important role in recognition and classification tasks. However, most of the traditional saliency detection methods did not take into account the texture features. The present method (Fig. 5) employs Gabor filter [9] for texture characterization. The method has two parallel passes as shown in Fig. 6. In the first pass the true color RGB image is subjected to superpixel segmentation employing SLIC and DBSCAN methods as described in the previous section; whereas, in parallel pass, the colored image has first been converted to gray scale image, then the obtained gray scale image is subjected to a number of directional filters varying in scale and orientations. In the present work, 5 scales and 8 orientations of Gabor filters are used. As a result, the same number of Gabor response images are obtained as the number of directional filters used. In the due course of feature extraction, all the regions segmented from the first pass are taken into consideration and response images from the second pass are used to compute *regional mean* and *variance vectors* for each region.

Algorithm 2 Proposed algorithm (Tex. sal. map).

1: **procedure** TEX–SALMAP(**Input** Image $I_{rgb}(x, y)$)
2: $I_{superpix}(x, y) = DBSCAN(SLIC(I_{rgb}(x, y)))$
3: $I_{gs}(x, y) = grayscale(I_{rgb}(x, y))$
4: Call Gabor filter Bank
5: $K = m * n$
6:     **for** each filter $k \in K$ **do**
7:         $R_k(x, y) = I_{gs}(x, y) * h_k(x, y)$
8:     **end for**
9:     **for** each $s \in m$ **do**
10:         **for** each $i \in N$ and $r_i(x, y) \in I_{superpix}$ **do**
11:             **for** each $o \in n$ **do**
12:                 $\sigma_{sio} = \sigma_{sio} + r_{sio}\{r_i(x, y)\}$
13:             **end for**
14:             $\mu_{sio} = \frac{1}{n}\sigma_{sio}$
15:         **end for**
16:     **end for**
17: $\mu_i = \frac{1}{m}(\sum_{s=1}^{m} \sigma_{sio})$
18: $V_i = Avg(\sum_{i=1}^{N}(r_i - \mu_i))$
19: $TEX - SAL = \sum_{j \neq i} \frac{N_j}{N_i}[(\mu_i - \mu_j)^2 + (V_i - V_j)^2]$
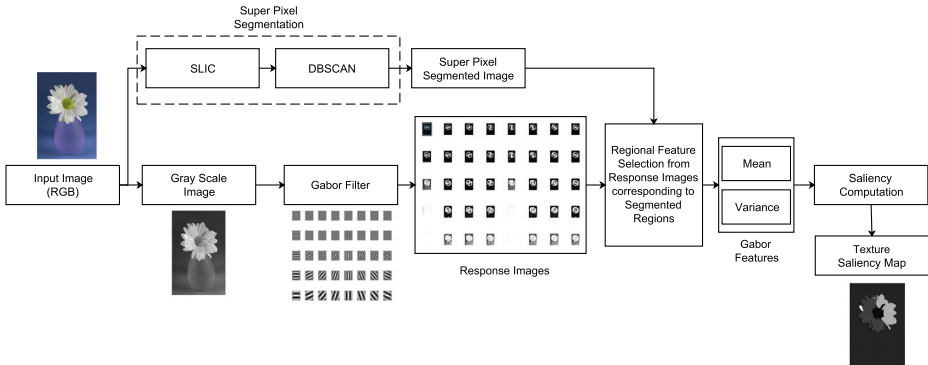20: **end procedure**

Fig. 5 Texture saliency map

For each scale, average of all the *regional mean* and *variance vectors* are computed. Thus, $N \times S$ regional mean and variance vectors are obtained. These vectors are further averaged to obtain N, number of *regional mean* and *variance vectors*. Thus, representing each region by a vector consisting of single mean and variance. Moreover, the features so computed are used as follows:

$$Mean(\mu_i)_{(s,o)} = \sum_{s=1}^{5} \sum_{o=1}^{8} x_i(s, o) \tag{8}$$

$$Variance(V_i)_{(s,o)} = \sum_{s=1}^{5} \sum_{o=1}^{8} y_i(s, o) \tag{9}$$

$$TextureSaliency = \sum_{j \neq i} \frac{N_j}{N_i} \left[(\mu_i - \mu_j)^2 + (V_i - V_j)^2 \right] \tag{10}$$

The color and texture saliency maps are combined by employing linear combination. Linear combination is used because, the corresponding weights can easily be adjusted on the basis of color and texture content with in the image.

$$S(r_j) = \alpha S_{col}(r_j) + \beta S_{tex}(r_j) \tag{11}$$

where α and β are the weights of color and texture saliencies respectively for region r_j.

6 Experimental results and discussions

In the present section, performance of the proposed algorithm has been assessed on the basis of its competence to predict visual saliency maps with reference to Ground Truths (GT). The GT images are provided with the used dataset. The proposed method has also been compared with the aforementioned six existing algorithms described in Section 2. Some results of these algorithms are given in Figs. 7 and 8.

In order to evaluate the performance of the proposed method, *MSRA 10K* dataset [30] has been used. As the proposed method is compared with six existing techniques: AC [1], graph based method (GB) [13], IG [2], Itti's method (IT) [19], MZ [32] and SR [16]. The
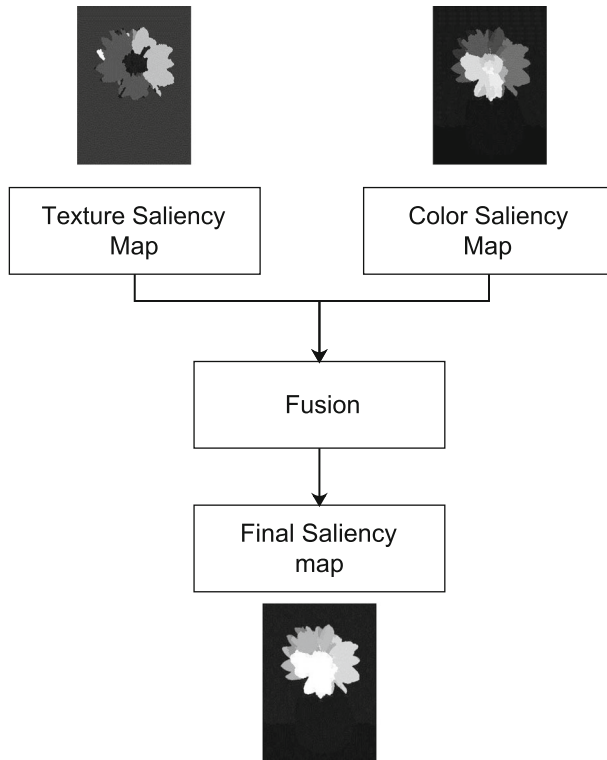
Fig. 6 Final saliency map

saliency maps of these methods are obtained from the author respective websites for the purpose of comparison.

6.1 Ground truth dataset: MSRA 10K

The MSRA 10K salient object database is a widely used dataset for salient object detection and segmentation. It consists of 10000 images and provides salient object annotations in terms of bounding boxes by 3-9 users. The corresponding ground truths are also provided with this dataset. The given images are of various sizes and in JPEG format.

6.2 Representative results and analysis

In Figs. 7 and 8, the saliency maps generated from the proposed method and other existing methods are shown. The assessment procedure has been divided into two parts; first the subjective assessment of the resultant saliency maps are performed, followed by an objective assessment of the corresponding images. The subjective assessment of these methods reveals that the saliency maps of the MZ [32] algorithm have high saliency values at the boundaries than the inner regions. Whereas, IT [19] is capable of detecting only fragments of the salient objects. AC [1] is found to be marking the salient object correctly but difficult to distinguish it from the background. GB [13] method is able to detect the salient object

Fig. 7 Visual Results of Saliency detection employing different methods. **a** Original Image, **b** AC **c** GB **d** IG **e** IT **f** MZ **g** SR and **h** the Proposed Method

completely in some images whereas partially in others. In case of SR [16] even the boundaries are partially detected. IG [2] provides somehow better results for all the images as it is detecting complete objects. However, the results of IG and all other given methods are outperformed by the proposed method.

Fig. 8 Visual Results of Saliency detection employing different methods. **a** Original Image, **b** AC **c** GB **d** IG **e** IT **f** MZ **g** SR and **h** the Proposed Method
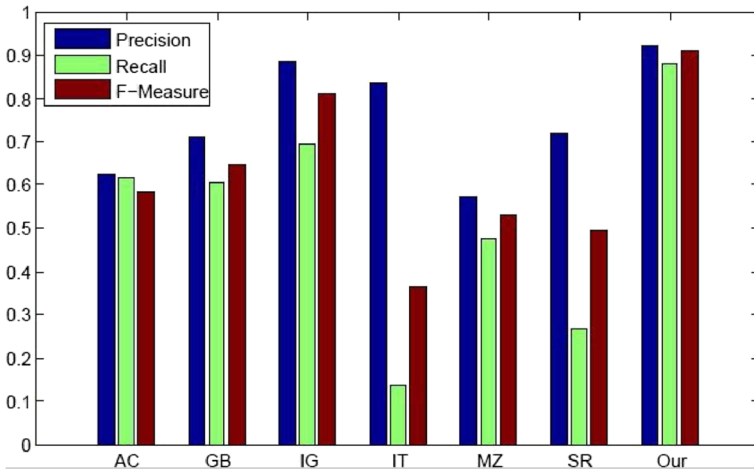
Fig. 9 Precision, recall and F-measure

The objective assesment includes the computation of Precision, Recall, F-measure and Mean Absolute Error (Figs. 9 and 10). In order to compute the first three measures i.e. Precision, Recall and F-measure, the saliency maps obtained from all the methods are thresholded using Otsu's [25] thresholding. The thresholding is carried out as follows: let the threshold value, for a particular image is T_v from Otsu's method [25]. The saliency regions where the pixels intensities are above the prescribed threshold T_v are regarded as salient regions, whereas the values lesser than T_v are considered as non-salient one's. In this way, images are converted to binary. In the binary image, white pixels show the salient objects whereas the black pixels show the non-salient regions. Now, the Precision is defined as the ratio of pixels that are classified accurately whereas Recall is described in terms of the number of
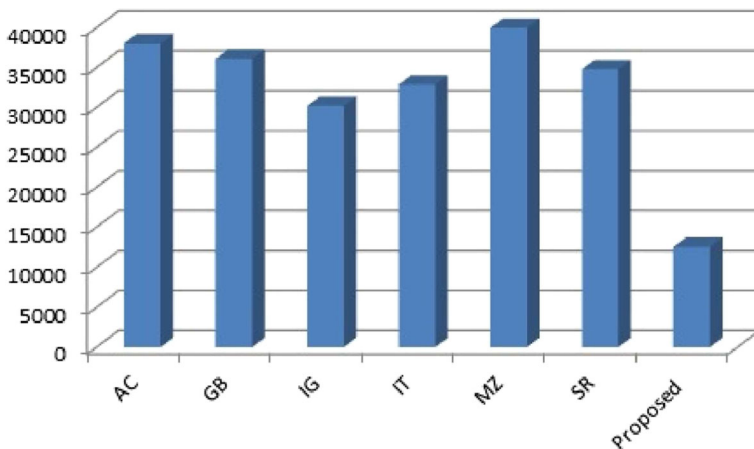

Fig. 10 Mean absolute error (MAE)

detected salient pixels to the total number of salient pixels. F-measure score combines both Precision and Recall. The three measures can be written mathematically as:

$$Precision = \frac{TP}{SM} \tag{12}$$

$$Recall = \frac{TP}{GT} \tag{13}$$

$$F - measure = \frac{(1 + \beta^2) precision \times recall}{\beta^2 \times precision + recall} \tag{14}$$

where TP stands for true positive, SM for saliency map and GT for ground truth. The value of β^2 is taken as 0.3. The precision, recall and F-measure values are shown in Fig. 9. From these graphs, it can easily be concluded that our method outperforms the other methods.

The evaluation measures used above are based upon overlapping strategy. These measures did not consider true negative saliency assignments; hence, may favor those methods that assigns high saliency to salient pixels but are not able to detect non-salient regions successfully. For a more comprehensive evaluation, authors have employed Mean Absolute Error (MAE). It takes into account the continuous saliency map in the same form, instead of changing it to binary and the binary ground-truth both. The two are used after a normalization in the range[0, 1]. Mathematically, it can be written as

$$MAE = \frac{1}{r \times c} \sum_{x=1}^{r} \sum_{y=1}^{c} \|S(x, y) - GT(x, y)\| \tag{15}$$

where S represents saliency map, GT is binary Ground-truth, r and c represents respective row and column size of the images.

As depicted in Fig. 10, MZ [32] is showing the highest error, then AC [1], GB [13], SR [16], IT [19] and IG [2]. The proposed method is showing the lowest value of MAE which is several times less than the existing methods. As evident from the subjective and objective assesments, the proposed method outperforms the existing one's because it has addressed the problem of saliency computation as two individual and separate subproblems i.e. color saliency computation and texture saliency computation. The two saliency maps are combined to obtain the final saliency map. Thus, the constituting saliency maps are complementing each other whereas the existing techniques have completely ignored the texture saliency computations. The computational complexity for color saliency map is $O(N^2)$ whereas for texture saliency map it is $O(m^2n^2)$ where m is the number of orientations and n is the number of scales for Gabor filter bank.

7 Conclusion

The paper presents a novel method for salient object detection and uses both color and texture cues. It forms two independent complementary maps for color and texture saliencies respectively. The method employs superpixel segmentation and implemented it using SLIC. The segments resulting from SLIC are tiny and large in numbers. Hence, these tiny regions are clustered using DBSCAN. The method incorporates two levels of color quantization employing minimum variance quantization and histogram before the computation of color saliency. The quantizations reduced the overall computation. Moreover, in saliency computations, centre surround principle is employed. In addition to this, texture features are

extracted using Gabor filters. Finally, the two complementary maps: color and texture are combined to obtain the final saliency map. The experimental results proves the efficacy of the proposed method and outperforms the other state of the art algorithms.

References

1. Achanta R, Estrada F, Wils P, Süsstrunk S (2008) Salient region detection and segmentation. In: International conference on computer vision systems. Springer, pp 66–75
2. Achanta R, Hemami S, Estrada F, Susstrunk S (2009) Frequency-tuned salient region detection. In: IEEE conference on computer vision and pattern recognition, 2009. cvpr 2009, IEEE, pp 1597–1604
3. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S (2012) Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans Pattern Anal Mach Intell 34(11):2274–2282
4. Chang KY, Liu TL, Chen HT, Lai SH (2011) Fusing generic objectness and visual saliency for salient object detection. In: 2011 IEEE international conference on computer vision (ICCV). IEEE, pp 914–921
5. Chang KY, Liu TL, Lai SH (2011) From co-saliency to co-segmentation: an efficient and fully unsupervised energy minimization model. In: 2011 IEEE conference on computer vision and pattern recognition (cvpr), IEEE, pp 2129–2136
6. Chen T, Cheng MM, Tan P, Shamir A, Hu SM (2009) Sketch2photo: Internet image montage. ACM Trans Graph (TOG) 28(5):124
7. Chen Zh, Liu Y, Sheng B, Jn Liang, Zhang J, Yb Yuan (2016) Image saliency detection using gabor texture cues. Multimed Tools Appl 75(24):16,943–16,958
8. Cheng MM, Mitra NJ, Huang X, Torr PH, Hu SM (2015) Global contrast based salient region detection. IEEE Trans Pattern Anal Mach Intell 37(3):569–582
9. Clausi DA, Jernigan ME (2000) Designing gabor filters for optimal texture separability. Pattern Recog 33(11):1835–1849
10. Goferman S, Zelnik-Manor L, Tal A (2012) Context-aware saliency detection. IEEE Trans Pattern Anal Mach Intell 34(10):1915–1926
11. Guo C, Zhang L (2010) A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. IEEE Trans Image Process 19(1):185–198
12. Han J, Ngan KN, Li M, Zhang HJ (2006) Unsupervised extraction of visual attention objects in color images. IEEE Trans Circuits Syst Video Technol 16(1):141–145
13. Harel J, Koch C, Perona P (2007) Graph-based visual saliency. In: Advances in neural information processing systems, pp 545–552
14. Hiremath P, Pujari J (2008) Content based image retrieval using color boosted salient points and shape features of an image. Int J Image Process 2(1):10–17
15. Hou X, Zhang L (2007) Saliency detection: A spectral residual approach. In: IEEE conference on computer vision and pattern recognition, 2007. CVPR'07. IEEE, pp 1–8
16. Hou X, Harel J, Koch C (2012) Image signature: highlighting sparse salient regions. IEEE Trans Pattern Anal Mach Intell 34(1):194–201
17. Hu Y, Rajan D, Chia LT (2005) Robust subspace analysis for detecting visual attention regions in images. In: Proceedings of the 13th annual ACM international conference on multimedia, pp 716–724
18. Huang H, Zhang L, Fu TN (2010) Video painting via motion layer manipulation. In: Computer graphics forum, vol 29. Wiley Online Library, pp 2055–2064
19. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259
20. Ji QG, Fang ZD, Xie ZH, Lu ZM (2013) Video abstraction based on the visual attention model and online clustering. Signal Process Image Commun 28(3):241–253
21. Jiang P, Ling H, Yu J, Peng J (2013) Salient region detection by ufo: Uniqueness, focusness and objectness. In: 2013 IEEE international conference on computer vision (ICCV). IEEE, pp 1976–1983
22. Jonathan H, Christof K, Pietro P (2006) Graph-based visual saliency. In: Proceedings of the 20th annual conference on neural information processing systems, pp 545–552
23. Judd T, Ehinger K, Durand F, Torralba A (2009) Learning to predict where humans look. In: 2009 IEEE 12th international conference on computer vision. IEEE, pp 2106–2113

24. Klein DA, Frintrop S (2011) Center-surround divergence of feature statistics for salient object detection. In: 2011 IEEE international conference on computer vision (ICCV). IEEE, pp 2214–2219
25. Kurita T, Otsu N, Abdelmalek N (1992) Maximum likelihood thresholding based on population mixture models. Pattern Recogn 25(10):1231–1240
26. Li A, She X, Sun Q (2013) Color image quality assessment combining saliency and fsim. In: 5th international conference on digital image processing (ICDIP 2013), international society for optics and photonics, vol 8878, p 88780I
27. Li J, Levine MD, An X, Xu X, He H (2013) Visual saliency based on scale-space analysis in the frequency domain. IEEE Trans Pattern Anal Mach Intell 35(4):996–1010
28. Li X, Li Y, Shen C, Dick A, Van Den Hengel A (2013) Contextual hypergraph modeling for salient object detection. In: 2013 IEEE international conference on computer vision (ICCV). IEEE, pp 3328–3335
29. Liu F, Gleicher M (2006) Region enhanced scale-invariant saliency detection. In: 2006 IEEE international conference on multimedia and expo. IEEE, pp 1477–1480
30. Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum HY (2011) Learning to detect a salient object. IEEE Trans Pattern Anal Mach Intell 33(2):353–367
31. Lou J, Ren M, Wang H (2014) Regional principal color based saliency detection. PloS one 9(11):e112,475
32. Ma YF, Zhang HJ (2003) Contrast-based image attention analysis by using fuzzy growing. In: Proceedings of the 11th ACM international conference on multimedia. ACM, pp 374–381
33. Meger D, Forssen PE, Lai K, Helmer S, McCann S, Southey T, Baumann M, Little JJ, Lowe DG (2008) Curious george: an attentive semantic robot. Robot Auton Syst 56(6):503–511
34. Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381(6583):607
35. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 9(1):62–66
36. Perazzi F, Krahenbuhl P, Pritch Y, Hornung A (2012) Saliency filters: contrast based filtering for salient region detection. In: 2012 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 733–740
37. Portilla J, Navarro R, Nestares O, Tabernero A (1996) Texture synthesis-by-analysis method based on a multiscale early-vision model. Opt Eng 35(8):2403–2417
38. Ren YF, Mu ZC (2014) Salient object detection based on global contrast on texture and color. In: 2014 international conference on machine learning and cybernetics (ICMLC), vol 1. IEEE, pp 7–12
39. Rosin PL (2009) A simple method for detecting salient regions. Pattern Recogn 42(11):2363–2371
40. Rutishauser U, Walther D, Koch C, Perona P (2004) Is bottom-up attention useful for object recognition? In: Proceedings of the 2004 IEEE computer society conference on computer vision and pattern recognition, 2004, CVPR 2004, vol 2. IEEE, pp II–II
41. Sharma G, Jurie F, Schmid C (2012) Discriminative spatial saliency for image classification. In: 2012 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 3506–3513
42. Shen X, Wu Y (2012) A unified approach to salient object detection via low rank matrix recovery. In: 2012 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 853–860
43. Siagian C, Itti L (2007) Rapid biologically-inspired scene classification using features shared with visual attention. IEEE Trans Pattern Anal Mach Intell 29(2):300–312
44. Tatler BW (2007) The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. J Vis 7(14):4–4
45. Tian H, Fang Y, Zhao Y, Lin W, Ni R, Zhu Z (2014) Salient region detection by fusing bottom-up and top-down features extracted from a single image. IEEE Trans Image Process 23(10):4389–4398
46. Valenti R, Sebe N, Gevers T (2009) Image saliency by isocentric curvedness and color. In: 2009 IEEE 12th international conference on computer vision. IEEE, pp 2185–2192
47. Wang P, Wang J, Zeng G, Feng J, Zha H, Li S (2012) Salient object detection for searched web images via global saliency. In: 2012 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 3194–3201
48. Xia C, Qi F, Shi G, Wang P (2015) Nonlocal center–surround reconstruction-based bottom-up saliency estimation. Pattern Recog 48(4):1337–1348
49. Yan Q, Xu L, Shi J, Jia J (2013) Hierarchical saliency detection. In: 2013 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 1155–1162
50. Yu Z, Wong HS (2007) A rule based technique for extraction of visual attention regions based on real-time clustering. IEEE Trans Multimed 9(4):766–784
51. Zhai Y, Shah M (2006) Visual attention detection in video sequences using spatiotemporal cues. In: Proceedings of the 14th ACM international conference on multimedia. ACM, pp 815–824

52. Zhang L, Yang L, Luo T (2016) Unified saliency detection model using color and texture features. PloS One 11(2):e0149,328
53. Zhou L, Yang Z, Yuan Q, Zhou Z, Hu D (2015) Salient region detection via integrating diffusion-based compactness and local contrast. IEEE Trans Image Process 24(11):3308–3320

Mudassir Rafi did his B.Sc with Honours in Geology from Aligarh Muslim University, Aligarh, India and M.C.A. from Jamia Millia Islamia, New Delhi, India in 2006 and 2010 respectively. He is currently pursuing Ph.D. in image processing from Indian Institute of Technology (ISM), Dhanbad, India. His research interest includes Salient Object detection, Texture analysis, Texture and Image segmentation.

Susanta Mukhopadhyay did his B.Sc. with Honours in Physics from Presidency College, Calcutta, B.Tech and M.Tech in Radiophysics and Electronics from the University of Calcutta and Ph.D. in Image Processing from the Indian Statistical Institute, Calcutta in 1988, 1992 and 2003 respectively. During 2001–2003 and 2004–2007 he worked at the Sanford Burnham Prebys Medical Discovery Institute, La Jolla, California, USA as postdoctoral researcher and Nanyang Technological University, Singapore, as research fellow respectively. His research area and interest include image processing, fMRI, image compression, image encryption and image watermarking. He is currently working as Associate Professor in the Department of Computer Science and Engineering, IIT, Dhanbad, India.